SUSANA SOBRINO.

CAPÍTULO I

INTRODUCCIÓN A LA TEORÍA DEL MUESTREO

1.1. Introducción y conceptos básicos

1.2. Fases de una investigación por muestreo

1.2.1. Planificación

1.2.2. El trabajo de campo

1.2.3. Explotación de los datos

1.3. Errores en el muestreo

1.4. Tamaño de la muestra

1.5. Principales tipos de muestreo

1.5.1. Muestreos no probabilísticos

1.5.2. Muestreos probabilísticos

TEXTO: Fotocopias de Apuntes

BIBLIOGRAFÍA COMPLEMENTARIA:

MILLS, R. L.: *Estadística para economía y administración*, Madrid, McGraw-Hill; capítulo 5, págs.137-140.

1.1. Introducción y conceptos básicos

El conocimiento de la realidad implica, generalmente, el estudio y análisis de ciertas características de una población, para lo cual es necesario obtener la información de todas las "unidades" de la población, es decir, realizar un *censo*. Sin embargo, la realización de censos suele resultar costosa, siendo además necesario mucho tiempo y su elaboración complicada, por lo que para obtener esta información tendremos que recurrir, en la mayoría de las ocasiones, a métodos que nos reduzcan el coste, la duración y la complicación del procedimiento.

Esto lo podemos hacer mediante la obtención de una muestra que nos refleje, lo más fielmente posible, el conjunto de la población objeto de estudio, aunque la realmente estudiada (sobre la que se obtiene la muestra) no coincida, por lo general, con la población que queremos estudiar, debido a la existencia de duplicidades, omisiones, inaccesibilidad de los datos, etc.

Para seleccionar la muestra, hay que establecer previamente cuáles son las *unidades de muestreo*, es decir, cuáles son los elementos sobre los que se realiza el muestreo. Esto algunas veces no presenta mayor problema, pues la unidad muestral es obvia, como en el caso, por ejemplo, de que queramos saber la opinión de los alumnos de un colegio sobre las Pruebas de Acceso a la Universidad —está claro que la *unidad muestral* será el *alumno*—; pero en ocasiones, esto no queda tan claro, así, si la población objeto de

4º) Organización del trabajo de campo

a) Selección previa de entrevistadores

b) Entrenamiento de entrevistadores

c) Tipos de entrevista

- Telefónica
- Postal
- Personal
- De grupo

5º) Tratamiento estadístico de los resultados de la encuesta

a) Inspección

b) Depuración

c) Codificación

d) Tabulación

e) Análisis de resultados

1.3. Errores en el muestreo

Toda muestra pretende ser representativa de la población que procede, si el diseño de la misma ha sido realizado con todas garantías. Sin embargo, no siempre el diseño muestral es el adecuado y aunque lo sea, siempre existen unas fuentes potenciales de errores que será necesario identificar.

Vamos a distinguir entre *dos tipos de errores* que se pueden cometer en el muestreo.

A) El error no muestral o sistemático

Un error que es cometido con frecuencia, es el denominado *error de cobertura*, que es el que se comete al no coincidir la población objeto de estudio con la población marco sobre la que se está obteniendo la muestra, debido a la existencia de omisiones, duplicidades, etc.

Otro tipo de error no muestral que se suele cometer, es el que se da por la *falta de respuesta* de algunos elementos seleccionados en la muestra y que producen, por una parte, la disminución de la precisión como consecuencia de que el tamaño de la muestra disminuye, y por otra, un sesgo en el resultado.

La utilización de instrumentos de medición incorrectos, cuestionarios mal definidos y técnicas de entrevistas inadecuadas suelen ser errores que conducen a que las opiniones de los entrevistados no sean las que de verdad hubiesen dado.

B) Errores muestrales

Son los errores que se cometen como consecuencia de la utilización de una muestra.

Este tipo de errores representan las diferencias que pueden darse entre el *estadístico muestral* y su *parámetro poblacional* correspondiente. El *estadístico* o *estimador* es el valor en la muestra de lo que estamos estimando. Por ejemplo: si estamos estimando la media de las alturas de los alumnos de una clase en base a una muestra de 5 alumnos, la media de las alturas de esos 5 alumnos será el estadístico media

generalizar los resultados obtenidos.

¿De qué depende el tamaño de la muestra? El principal factor a tener en cuenta es el nivel de precisión requerido o, lo que es lo mismo, el error de muestreo que nos hayamos fijado; pero, además, existen otros factores que se deben tener en cuenta, como:

- **La varianza poblacional**

Si tenemos en cuenta que cuanto más homogénea es la población objeto de estudio, su varianza es menor y, por lo tanto, el número de elementos necesarios en la muestra será menor que si la población es heterogénea.

Por lo que es necesario conocer a priori la *homogeneidad* o *heterogeneidad* de la población, ya que necesitamos tener un conocimiento aproximado de la varianza para decidir sobre el tamaño de la muestra, para lo cual se suele recurrir a la experiencia o a estudios piloto previos a la investigación.

- **El nivel de confianza**

Se elige subjetivamente por el investigador y nos da la probabilidad de acertar, es decir, la probabilidad de que si no hay sesgos, la estimación realizada sea la verdadera. En realidad, se trata de una acotación de la función de probabilidad o de la función de densidad.

- **El tipo de muestreo**

Como veremos en el siguiente apartado, se pueden realizar distintos tipos de muestreo en función del objetivo de la investigación, de las características de la población o del coste que nos podamos permitir. Pues bien, esto tendrá su importancia en el cálculo del tamaño de la muestra, ya que, como veremos, la precisión varía según se realice un tipo u otro de muestreo.

1.5. Tipos de muestreo

Hasta ahora hemos planteado como objetivo fundamental de la teoría del muestreo, la obtención de una muestra representativa de la población. No obstante, convendría precisar esta afirmación general considerando el factor económico y estableciendo como objetivo final el de *la selección de una muestra que, al mínimo coste, nos permita obtener la mejor representación de la población.*

En consecuencia, la teoría del muestreo ha desarrollado un amplio abanico de diferentes tipos de muestreo, en función del mayor o menor hincapié que se quiera realizar sobre los dos criterios básicos de elección:

- Representatividad
- Coste.

Obviamente, ambos criterios se contraponen. La máxima representatividad podría alcanzarse, en principio, mediante la realización de un censo o estudio exhaustivo de todas las

- **Muestreo por cuotas**

Normalmente se utiliza en combinación con otros tipos de muestreo y en la etapa final de la recogida de datos o labor de campo. Supone que el entrevistador, dada una cuota preestablecida de observaciones a obtener, elige él mismo los elementos a encuestar (por ejemplo, la encuesta fotográfica que aparece diariamente en las páginas del *Diario Vasco*). Hay que llenar una cuota de observaciones, pero no hay determinación de a quién observar.

1.5.2. Muestreos probabilísticos

Presentaremos a continuación los principales tipos de muestreos aleatorios o probabilísticos. La variedad de los mismos se explica por la diferente forma y énfasis puesto en la consecución de los dos objetivos contrapuestos:

- Seleccionar una muestra que represente bien a la población. En este sentido será necesario siempre, por una parte, un buen conocimiento del problema objeto de estudio, del comportamiento de la característica de interés, en suma, de la naturaleza de la población a muestrear, y de otra, al mismo tiempo, un buen conocimiento de las características, ventajas e inconvenientes de los diversos tipos de muestreo disponibles con el fin de garantizar la elección del tipo de muestreo más adecuado, el que nos ''asegure'' la obtención de una muestra representativa.

- Disminuir los costes asociados al muestreo.

- **Muestreo aleatorio simple**

Es el más elemental y básico de todos los tipos de muestreo aleatorios existentes. Para seleccionar una muestra aleatoria simple basta con que se cumpla el requisito de que cada elemento de la población tenga la misma probabilidad de ser seleccionado para entrar a formar parte de la muestra. En consecuencia, la probabilidad que tiene cada muestra de tamaño ''n'' de ser seleccionada de la población es la misma.

En la práctica, dada precisamente su extremada sencillez, es raramente utilizado. Hay que tener en cuenta que cada elemento de la población puede ser elegido. En consecuencia, sólo si las observaciones de la característica de interés en la población se comportan de forma homogénea, similar y uniforme (por ejemplo, edades de los alumnos de 3º de la E.S.T.E.) puede esperarse obtener, mediante un muestreo aleatorio simple, una muestra bastante representativa. El factor económico, por otra parte, constituye un serio inconveniente en el caso de observaciones heterogéneas, ya que para conseguir una buena representatividad de la muestra habría que seleccionar muchas observaciones.

En suma, este tipo de muestreo tiene un interés de conocimiento casi estrictamente pedagógico, en tanto en cuanto una vez estudiado permite comprender más rápidamente los diversos tipos de muestreos probabilísticos.

- **Muestreo aleatorio sistemático**

Constituye un caso particular del muestreo aleatorio simple y se aplica para agilizar la selección de las

observaciones de la muestra cuando el tamaño de la población es elevado.

Supongamos una población compuesta por N = 1000 individuos y que de ella se quisiera seleccionar una muestra de n = 200 individuos. Si aplicásemos un muestreo aleatorio simple tendríamos que seleccionar uno por uno a los 200 integrantes de la muestra, cuidando que en cada selección se mantuviese la equiprobabilidad. El proceso a seguir, en cambio, si seleccionásemos la muestra mediante un muestreo aleatorio sistemático sería el siguiente:

a) Dividir el marco en tantos bloques iguales como observaciones vaya a contener la muestra. En nuestro caso, 1000/200 = 5, es decir, formaríamos en el marco 200 bloques de 5 elementos cada uno (1-5, 6-10, 11-15, ..., 996-1000).

b) Seleccionar un número al azar (de forma equiprobable) en el primer bloque (por ejemplo, en el bloque 1-5 resulta elegido al azar el número 3).

c) A partir de ahí, se seleccionará la muestra total en el resto de los bloques, tomando el elemento correspondiente en el orden al del seleccionado en el primer bloque (el 3, el 8, el 13, el 18, el 23, ..., el 998).

Este método de selección asegura que la probabilidad de elección es la misma para todos los elementos de la población (propiedad del muestreo aleatorio simple), siendo ésta n/N (en nuestro caso 200/1000 = 1/5), a diferencia del muestreo aleatorio simple en que era de 1/N (en nuestro caso 1/1000).

Las *ventajas* del muestreo aleatorio sistemático frente al muestreo aleatorio simple son:

- Su sencillez y comodidad de aplicación.

- Permite obtener una muestra distribuida a lo largo del marco, pudiendo resultar en consecuencia más representativa que si la selección de cada elemento hubiese sido individual, como plantea el muestreo aleatorio simple, donde por ejemplo, el azar pudiese haber hecho que los individuos seleccionados se concentraran en una zona determinada del marco.

Las *desventajas* se resumen, fundamentalmente, en la posible existencia de sesgos muy sutiles a consecuencia del método de selección utilizado. Por ejemplo, planteemos la selección a lo largo del año de una muestra de días de trabajo, por lo que se construyen bloques de una semana. Este método nos llevaría a tomar, por ejemplo, una muestra exclusivamente de los lunes, lo cual podría ser generador de sesgos en las conclusiones del análisis.

- **Muestreo aleatorio estratificado**

El interés de su aplicación surge cuando nos encontramos ante una población de observaciones que tienen un comportamiento heterogéneo ante la característica de interés considerada. Las observaciones son muy diferentes entre sí y existe alguna causa o causas externas que explican esa heterogeneidad.

Para ilustrar esta situación con un ejemplo aclaratorio, supongamos que una determinada empresa de ámbito estatal y con una elevada plantilla de vendedores considerase para el año próximo una modificación de los actuales sistemas

de comisiones, por lo que, entre otras informaciones, sería interesante disponer de una cifra estimada de las ventas medias por vendedor para el siguiente ejercicio. Las limitaciones de tiempo y dinero son tales que se plantea la realización de un muestreo con esta finalidad. Sin embargo, la empresa es consciente de la enorme diversidad (heterogeneidad) de la población a estudiar (cifras de ventas por vendedor en el año próximo), explicable principalmente por la diversidad de productos que atienden los distintos vendedores.

Supongamos que sólo se venden 3 tipos básicos de productos y que cada vendedor sólo vende un determinado tipo de producto. Gráficamente, la población objeto de estudio tendría la siguiente naturaleza:

Estrato 1: E1	Estrato 2: E2	Estrato 3: E3
Cifra de ventas por vendedor del producto 1	Cifra de ventas por vendedor del producto 2	Cifra de ventas por vendedor del producto 3
N1 = Tamaño de E1	N2 = Tamaño de E2	N3 = Tamaño de E3

$$N = N1 + N2 + N3$$

Si ante esta población planteamos la selección de una muestra por muestreo aleatorio simple, el azar no nos asegura que la muestra elegida vaya a contener vendedores de los tres tipos de productos. Podría suceder que sólo se seleccionasen

mayoritariamente vendedores de un determinado tipo de producto, alterándose totalmente la representatividad de la muestra y, por consiguiente, la fiabilidad de las conclusiones del análisis. Si por el contrario enfocamos el problema mediante el empleo de un muestreo aleatorio estratificado, el proceso a seguir sería el siguiente:

1º) Definición de la variable o variables de estratificación.

Es el aspecto más importante de todo el proceso, ya que dependiendo de la variable que se defina resultarán unos u otros estratos. En nuestro ejemplo, la variable de estratificación que explica principalmente la heterogeneidad de la población es el tipo de producto vendido. Normalmente la selección de esta variable se hace por juicios.

2º) Construcción de los marcos de cada estrato.

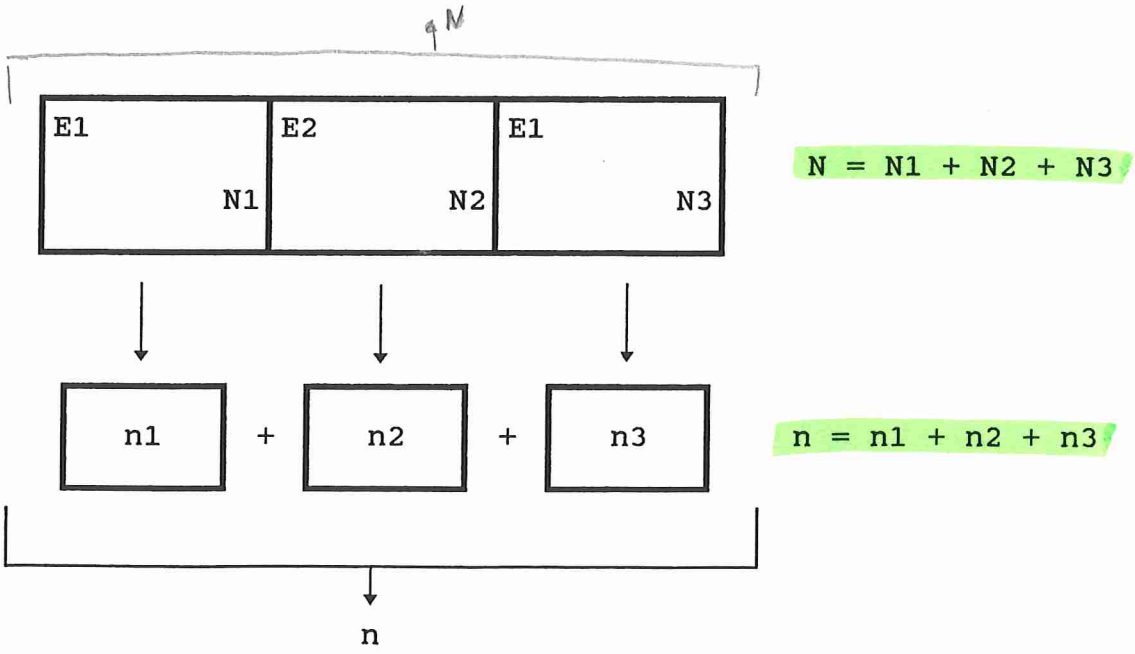
Una vez establecido el tipo de producto como variable de estratificación, procederíamos a clasificar a nuestros vendedores en 3 grandes listas o marcos por tipo de producto vendido.

El objetivo de estos dos pasos, y en suma de todo proceso de estratificación, no es otro que el de *construir estratos homogéneos dentro de sí y heterogéneos entre sí*. Si el tipo de producto fuese una buena variable de estratificación, habríamos conseguido definir 3 estratos o

subpoblaciones dentro de cada una de las cuales sólo encontraríamos observaciones muy similares, homogéneas, con poca dispersión en sus valores (cifras de ventas por vendedor de un determinado producto), mientras que simultáneamente, los estratos serían muy diferentes, heterogéneos unos de otros.

3º) Selección de una muestra en cada estrato.

Se procedería a continuación a la selección al azar por muestreo aleatorio simple (cada elemento del estrato tendría la misma probabilidad de resultar elegido en la muestra de dicho estrato) de una muestra en cada uno de los estratos. Se conformaría así la muestra total como suma de las muestras tomadas en cada uno de los estratos.



El problema que se plantea entonces es el de distribuir el tamaño total de la muestra entre los diferentes estratos, es decir, cómo repartir n entre n1, n2 y n3. Este proceso denominado *afijación*, puede llevarse a cabo de dos formas:

- *Proporcionalmente*: El tamaño de la muestra total se distribuye entre los diversos estratos en función del tamaño de cada estrato con respecto a la población.
- Óptimamente: El tamaño de la muestra total se distribuye entre los diversos estratos en función de la dispersión de las observaciones contenidas en cada estrato. Así, podrá haber estratos grandes con muy poca dispersión del que elegiríamos una muestra pequeña, y estratos pequeños con mucha dispersión del que elegiríamos muestras más grandes.

Una vez explicado básicamente el proceso de aplicación del muestreo aleatorio estratificado se puede comprender fácilmente su ventaja frente al muestreo aleatorio simple en el caso (muy habitual en la práctica) de que la población objeto de estudio tenga naturaleza heterogénea, ya que nos asegura, si la estratificación realizada es buena, que nuestra muestra sea una buena representación de la población, una buena "fotografía", fiel reflejo de las características de la población.

Dicho en otras palabras, para conseguir una buena muestra en este caso mediante un muestreo aleatorio simple sería preciso incrementar mucho el tamaño total de la

$$N = N1 + N2 + N3 + N4 + N5 + N6$$

Conglomerado 1: C1 Cifra de ventas por vendedor en la provincia 1 N1 = Tamaño de C1	Conglomerado 2: C2 Cifra de ventas por vendedor en la provincia 2 N2 = Tamaño de C2	Conglomerado 3: C3 Cifra de ventas por vendedor en la provincia 3 N3 = Tamaño de C3
Conglomerado 4: C4 Cifra de ventas por vendedor en la provincia 4 N4 = Tamaño de C4	Conglomerado 5: C5 Cifra de ventas por vendedor en la provincia 5 N5 = Tamaño de C5	Conglomerado 6: C6 Cifra de ventas por vendedor en la provincia 6 N6 = Tamaño de C6

Conglomerado 5: C5 Cifra de ventas por vendedor en la provincia 5 N5 = Tamaño de C5	Conglomerado 3: C3 Cifra de ventas por vendedor en la provincia 3 N3 = Tamaño de C3

$n = N5 + N3$

n

Se trata de q. cada conglomerado sea una representación en proporción de la población, y los q. cada conglomerado no deben ser heterogéneo dentro de sí, y han de ser homogéneos entre sí.

descrito un *muestreo aleatorio por conglomerados monoetápico* o en una única etapa. No sería difícil de imaginar en nuestro supuesto en que la muestra formada por los conglomerados 5 y 3 fuese a su vez ''conglomerada'' en zonas más restringidas (por ejemplo, comarcas), de tal modo que una vez definidos los nuevos conglomerados se tomase a su vez una nueva muestra de conglomerados. Este proceso podría prolongarse a voluntad, dando lugar a los llamados *muestreos aleatorios conglomerados polietápicos*.

INTRODUCCIÓN AL MUESTREO: EJEMPLOS

EJEMPLO 1:

La población a encuestar es de 10.000 elementos. Supongamos que la muestra elegida es de 400. Si en 8 de ellos hemos obtenido que respondían a una determinada característica: ¿cuántos elementos de la población representa?; ¿cuál es la fracción de muestreo?

EJEMPLO 2:

Queremos hacer un estudio sobre el tiempo que una persona permanece desempleada en Guipúzcoa; para ello previamente nos piden que diseñemos una muestra representativa.

EJEMPLO 3:

Queremos hacer un estudio que implica realizar un muestreo en una ciudad tomando como unidad de muestreo la vivienda, encontrándose éstas muy esparcidas por barrios. Si queremos una muestra de 500 viviendas: diseñar el plan de muestreo.

EJEMPLO 4:

Se quiere realizar una encuesta sobre las características socioeconómicas y de forma de vida de un barrio obrero. Los datos que quiero obtener son los relativos a cada vivienda. ¿Cómo plantearíamos la encuesta?